Research Paper | Agriculture | E-ISSN No : 2454-9916 | Volume : 8 | Issue : 11 | Nov 2022

A STOCHASTIC PROCESS IN MODELING AND FORECASTING OF ONION PRODUCTION IN INDIA

Dr. Sameerabanu P[1], Dr. C. Sekhar[1]

[1]Nalanda College of Agriculture, M.R.Palayam, Tiruchirappalli, India.

ABSTRACT

This work focuses on modeling and analysis of crop yield over space and time. Specifically, the onion yield data set was used. This study on estimating the future yield of onions in major producing states in India. To achieve this, we applied time series on onion yield data recorded from 1978 to 2020, as per availability from the website of Ministry of Agriculture, Government of India. By using SPSS software, The data are analyzed using the autoregressive integrated moving average (ARIMA) model to best fit the model. Selected best models were used to estimate onion yield. The selection of a suitable model requires determining the efficiency of different models in predicting future outcomes and selecting the most suitable model for the prediction work. To develop a suitable forecast ARIMA model for agricultural data. To study the predictive ability of the univariate ARIMA model to suggest an optimal model, the best predictive model was selected.

KEYWORDS: ARIMA, ACF, PACF, Onion forecast.

INTRODUCTION:

Farming sector faces several challenges from land preparation to harvesting and marketing of farm produce. The consumers of farm output though are healthy and wealthy; they are able to bargain to the lowest price for the output realized. The traders are having collusion and their association is strong enough to bargain from the farmers. But the farmers and the farmer organizations are weak in their association and cannot be unified to establish an organization to the fullest spirit to command price for their produce. Though there were few farmer organizations in our country, they cannot raise to the expected level in achieving or distributing the farm produce. During the time of harvest, supply will be excess and the demand will be less and hence proper storage and distribution is a must. State Governments took effort to procure the principal crop outputs particularly in respect of cereals like rice and cannot procure other farm produce timely and hence balancing the area under crop and its output production becomes much more important. For that, information on price availability, demand for that produce and expected price for the output by the farmers are to be provided in advance prior to the crop season. In this respect, forecasting is the tool that will help to predict the yield and price in advance. Forecasting refers to the practice of predicting what will happen in the future by taking into consideration events in the past and present. Basically, it is a decision-making tool that helps businesses cope with the impact of the future's uncertainty by examining historical data and trends. It is a planning tool that enables businesses to chart their next moves and create budgets that will hopefully cover whatever uncertainties may occur (CFI, 2022). This study is the one that aimed at forecasting the yield of onion in India.

METHODOLOGY:

This study aimed at forecasting the yield of Onion in India. For that the basis of forecasting is to be discussed to develop an overall idea. The first step in the process is developing the basis of the investigation and identifying where the business is currently positioned in the market.

Forecasting Methods:

Businesses choose between two basic methods when they want to predict what can possibly happen in the future, namely, qualitative and quantitative methods.

1. **Qualitative Method:** Otherwise known as the judgmental method, qualitative forecasting offers subjective results, as it is comprised of personal judgments by experts or forecasters. Forecasts are often biased because they are based on the expert's knowledge, intuition, and experience, and rarely on data, making the process non-mathematical.

   One example is when a person forecasts the outcome of a finals game in the NBA, which, of course, is based more on personal motivation and interest. The weakness of such a method is that it can be inaccurate.

2. **Quantitative Method:** The quantitative method of forecasting is a mathematical process, making it consistent and objective. It steers away from basing the results on opinion and intuition, instead utilizing large amounts of data and figures that are interpreted.

Features of Forecasting:

Here are some of the features of making a forecast:

- Forecasts are created to predict the future, making them important for planning.
- Forecasts are based on opinions, intuition, guesses, as well as on facts, figures, and other relevant data. All of the factors that go into creating a forecast reflect to some extent what happened with the business in the past and what is considered likely to occur in the future.
- Most businesses use the quantitative method, particularly in planning and budgeting activities

Collection of Data:

The secondary data was collected from the website of Ministry of Agriculture, Government of India. Onion production over a period of time was gathered from the above website and is analyzed using ARIMA models. By using SPSS software, the data was analyzed to fit the best model using an autoregressive integrated moving average (ARIMA) model. The selected best models were used to forecast the onion yield.

ARIMA Modeling:

In general, an ARIMA model is characterized by the notation ARIMA (p,d,q) where, p, d and q denote orders of autoregression integration (differencing) and moving average respectively. Time series is a linear function of past actual values and random shocks. For instance, given a time series process {Yi}, a first order auto-regressive process is denoted by ARIMA (1,0,0) or simply AR(1) and is given by

$$Y_i = \mu + \varphi_1 Y_{i-1} + \varepsilon_t$$

Stationary:

A stationary time series does not depend on the time when a particular point is observed. Each point in time has a value that is not dependant on another point in time, such as white noise. The plots below are some examples of stationary time series. Other examples may include cyclic data with non-consistent periods.

Differencing:

Differencing is a time series transformation that attempts to eliminate time-dependent factors from the time series such as trend and seasonality. There are different orders of differencing; the equation below shows the first-order difference. It is simply the difference between the current and previous observation:

$$y'_t = y_t - y_{t-1}$$

After the first-order difference, if the time series is still not stationary, differencing once more will give you the second-order differencing.

$$y''_t = y'_t - y'_{t-1}$$

$$= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$$

$$= y_t - 2y_{t-1} + y_{t-2}$$

The order of the differencing can be defined in the d parameter of the model.

Autoregressive Models:

An autoregressive (AR) model, defined as being the regression of it, is simply a

multiple linear regression having the previous time steps as parameters to the function. It usually performs better on stationary time series data.

$$y = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

The order of the AR model, p defines the number of previous time steps that are accounted for in the current observation. For example, AR(1) accounts for time step t-1 for the current observation, AR(2) accounts for time steps t-1 and t-2.

Moving Average Models:
Unlike AR models, Moving Average (MA) models predict the next step based on the errors of the previous steps.

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

or can be written as,

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t$$

The order of the MA model, q is similar to AR's. A larger q means that it takes into account of more number of previous time steps. ARIMA By combining auto regression, differencing and moving average, we get an ARIMA (p, d, q) model. p, d, q are respective parameters for AR, differencing and MA.

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Here, C is some constant + linear combination of the previous p + linear combination of the previous q error terms + the error this time (ε_t). y'_t is the differenced series. It needs to be integrated (opposite of differencing) to get the actual series.

RESULTS AND DISCUSSION:
Using the above models, the data outlined in Table 1 was analysed and the results are presented in the subsequent headings.

Table 1: Area and Productivity of Onion in India

S. No.	Year	Yield in Kgs./Hectare	Area in Million Hectares
1.	1978	10403	00.21
2.	1979	10232	00.24
3.	1980	9961	00.25
4.	1981	10562	00.25
5.	1982	10330	00.24
6.	1983	9982	00.27
7.	1984	11139	00.28
8.	1985	10202	00.28
9.	1986	9659	00.26
10.	1987	9857	00.27
11.	1988	10620	00.32
12.	1989	10176	00.30
13.	1990	10686	00.30
14.	1991	11088	00.32
15.	1992	10791	00.32
16.	1993	10902	00.37
17.	1994	10661	00.38
18.	1995	10316	00.40
19.	1996	10348	00.40
20.	1997	9091	00.40
21.	1998	11391	00.47
22.	1999	9932	00.49
23.	2000	10786	00.42
24.	2001	10686	00.45
25.	2002	9912	00.42
26.	2003	11784	00.50
27.	2004	11718	00.55
28.	2005	13118	00.66
29.	2006	12655	00.70
30.	2007	12974	00.70
31.	2008	16260	00.83
32.	2009	16079	00.76
33.	2010	14210	01.06
34.	2011	16109	01.09
35.	2012	15989	01.05
36.	2013	16120	01.20
37.	2014	16111	01.17
38.	2015	15857	01.32
39.	2016	17178	01.31
40.	2017	18103	01.28
41.	2018	18712	01.22
42.	2019	18875	01.43
43.	2020	18563	01.31

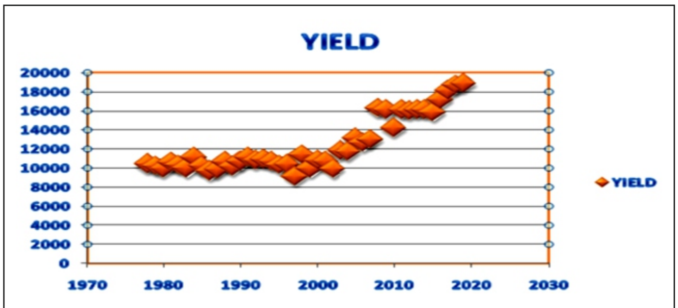

Figure: 1.1: The Magnitude of Yield of Onion in different Decades of Time

Modeling and Forecasting:
- To perform an actual forecast and assess quality of forecasts.

Identification:
- The graphs of ACF and PACF are drawn for all the observed variables, to the fitted models.

Sequence Plot:
The plotted diagram above indicated that over a period of time, the yield of onion per ha is found to be marginally increasing up to the year 2000 and later on the yield of onion and its magnitude found to be steadily increasing in an upward direction indicating that the onion production technology would have contributed to the higher yield over a period of time. Release of new hybrids in onion and its technology might be the reason for the upward swing. The involvement of horticulturist in developing onion hybrid and the role of extension functionaries in taking the technology to the farm level are the main reasons for adoption of technology and an upward yield realization.

Using this sequence diagram, the analysis has predicted the yield for the remaining years and is discussed suitably.

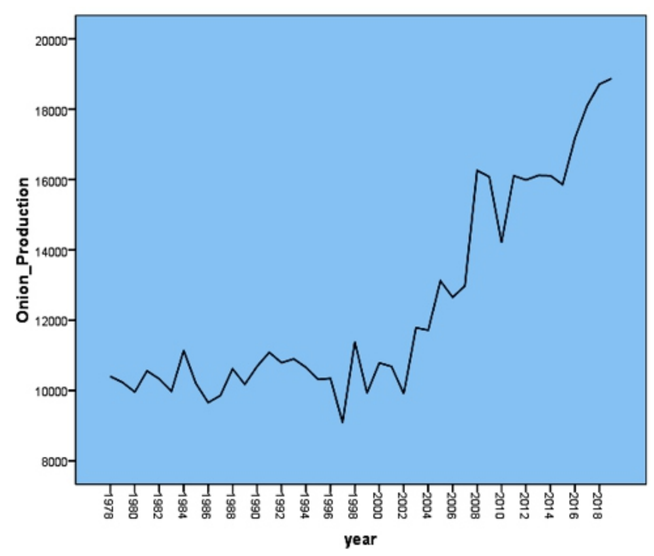

Figure: 1.2: Graph delineating Non Stationary in Onion Yield

Table 2: Autocorrelation and Partial Autocorrelation Correlograms for non stationary

Log	Autocorrelations	Std. Error	Partial Autocorrelations	Std. Error
1	.873	.149	.873	.154
2	.786	.147	.100	.154
3	.713	.145	.035	.154
4	.629	.143	-.067	.154
5	.580	.141	.092	.154

6	.492	.140	-.165	.154
7	.429	.138	.033	.154
8	.362	.136	-.068	.154
9	.268	.134	-.134	.154
10	.223	.132	.078	.154
11	.118	.130	-.246	.154
12	.008	.127	-.162	.154
13	-.023	.125	.220	.154
14	-.066	.123	.016	.154
15	-.121	.121	-.185	.154
16	-.150	.119	.136	.154

Correlogram Plot:

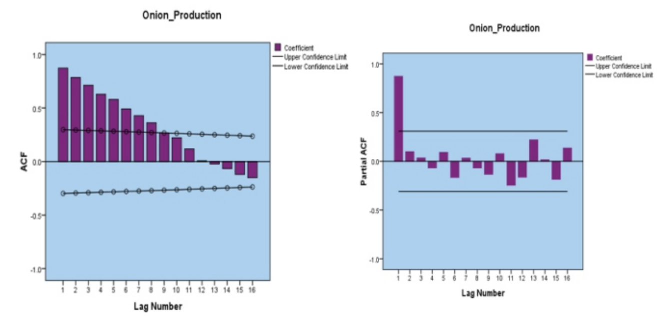

Figure: 2.1: Significance of Correlogram Plots

Since the above correlogram indicates non- stationary, maximum lag order 16.

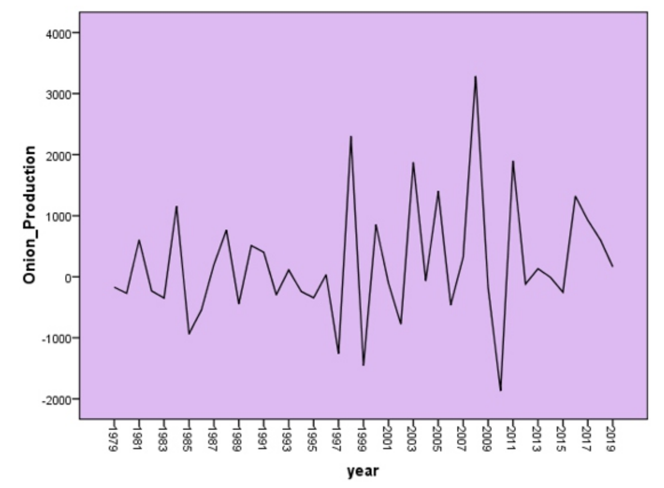

Figure: 2.2: The Sequence plot of Stationarity

Table: 2A: The Results of Autocorrelation and Partial Autocorrelations for stationary

Log	Autocorrelations	Std. Error	Partial Autocorrelations	Std. Error
1	-.370	.151	-.370	.156
2	-.083	.149	-.254	.156
3	.234	.147	.126	.156
4	-.132	.145	.000	.156
5	.244	.143	.306	.156
6	-.194	.141	-.054	.156
7	-.025	.139	-.068	.156
8	.140	.137	-.052	.156
9	-.131	.135	-.052	.156
10	.211	.133	.196	.156
11	-.051	.130	.171	.156
12	-.154	.128	-.075	.156
13	.158	.126	-.095	.156
14	-.018	.124	-.031	.156
15	-.005	.121	-.007	.156
16	-.018	.119	.044	.156

Correlogram Plots:

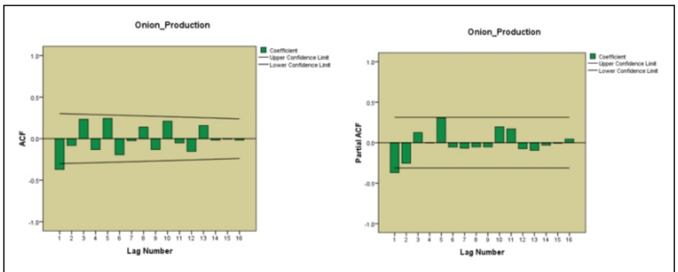

Figure: 2.3 : Correlogram Plots for Onion Yield

Since the above figure Correlogram indicates stationary, maximum lag order 16.

Forecast:

In this section, already defined models are used to forecast the Onion Yield from 1978-2020. Here also we perform the forecasting accuracy assessment at the end of this session, comparing forecast with the actual data.

Forecast Model Statistics:

Based on the forecasting performance, testing and estimating results, ARIMA (1, 1, 1). The residual ACF and PACF of ARIMA (1, 1, 1) provides a slightly better result. We can see that ACF and PACF of ARIMA (1, 1, 1) has a spike crossing the ±0. 1 boundary at lag 1, while nothing is crossing the boundary in ACF and PACF of ARIMA (1, 1, 1). It means that ARIMA (1, 1, 1) is slightly more stable), which probably provides better forecasting results.

Table: 3: Model statistics for Onion Yield

Model Statistics						
Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
OnionYield-Model_1	0	0.167	10.937	16	0.813	0

Table: 4: Forecast for Onion Yield with Upper Control and Lower Control Limits

Years to which Forecast was Made	Forecast Figures of Yield / Ha	Upper Control Limits	Lower Control Limits
2020	19006	20902	17109
2021	19225	21402	17048
2022	19434	21904	16963
2023	19644	22371	16916
2024	19854	22817	16891
2025	20063	23244	16882
2026	20273	23658	16888
2027	20483	24060	16906
2028	20693	24453	16933
2029	20903	24837	16969
2030	21112	25213	17012
2031	21322	25583	17061
2032	21532	25947	17117

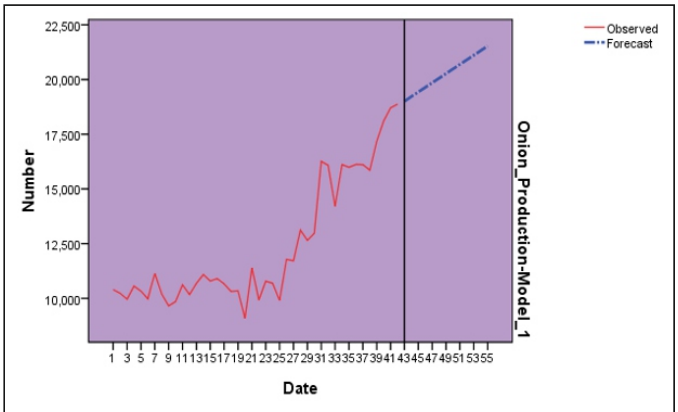

Figure: 4.1 Forecasted Yield of Onion and its Graphical Signature

Table: 5: Results of Goodness of Fit

Model Fit			
Fit Statistic	**Mean**	**Minimum**	**Maximum**
Stationary R-squared	0.167	0.167	0.167
R-squared	0.902	0.902	0.902
RMSE	937.277	937.277	937.277
MAPE	5.665	5.665	5.665
MaxAPE	18.973	18.973	18.973
MAE	688.584	688.584	688.584
MaxAE	3.085E3	3084.933	3084.933
Normalized BIC	13.958	13.958	13.958

The asterisks below indicate the best (that is, minimized) values of the respective information criteria,

- MAPE = Mean Absolute Percentage Error,
- RMSE = Root Mean Square Error and Normalized
- BIC = Normalized Bayesian Information criterion.

The above analysis tells us to identify the model as ARIMA (1,1,1) because all criteria significant. OLS estimates, observations 1978-2020, R squared = 90.20% our prediction is accurate.

Goodness of Fit:

Table 5 outlined the details of goodness of fit statistics for the Onion Yield data. R-squared represents an estimate of the proportion of the total variation in the series that is explained by the model. Here the R2 found to be 90.20 per cent indicates that the 90 per cent of the variation in the yield was explained by the independent variables included. Largest value (maximum value) indicates a more accurate prediction and it means that the model does an excellent job of explaining the observed variations in the series. Mean Absolute Percentage Error (MAPE) for the model is a measure of how much a dependent series varies from its model-predicted level. Root Mean Square Error (RMSE), indicates that the square root of mean square error is a measure of how much a dependent series varies from its model-level of prediction, expressed in the same units as the dependent series. This measure is useful for imagining a worst-case scenario for the forecast model.

SALIENT FINDINGS:

- ARIMA (1,1,1) was found to be the better applicable model from which the predictions are made for 2020-2032.
- ARIMA was used for the reasons of its capabilities of making predictions using a time series data with various kinds of pattern and with autocorrelations between the successive values in the time series.
- The study was also tested statistically and validated the residuals (forecast errors).
- The fitted ARIMA time series and residuals are seemed to be normally distributed with mean 0 and constant variance. Hence it can be concluded that the selected seasonal ARIMA (1,1,1) will provide an adequate model for interpreting and forecasting gold price in India.
- The ARIMA (1,1,1) model predicted indicates an increase in the Onion Yield for the years selected for the forecast study.

SUMMARY AND CONCLUSIONS:

In this research study, researcher analyzed and obtained the forecast of Onion Yield in India using ARIMA models. The result of the study conclude that ARIMA (1,1,1)model is the more appropriate model for forecasting Onion Yield in India. The forecast model and the forecast graph that the Onion Yield is rapidly increasing with the passage of year. Overall, we can see that ARIMA (1,1,1) provide a good fit for Onion Yield in India. Its gives a fairly accuracy forecasting. However, although forecast from 2020-2032 are within the 95 per cent interval, the graph shows that the blue line of actual data has gradually moving out of the confidence interval.

REFERENCES:

I. Box, G.E.P., Jenkin, G.M. and Reinse, G.C. 1994. The timeseries analysis, forecasting, and control. 3rd ed, Prentice-Hall, Englewood Cliffs, N.J.

II. Box, G.E.P. and Jenkin, G.M. 1976. Time-Series of Analysis, Forecasting, and Control. Sam Francisco, Holden-Day, California. the USA.

III. Darekar, A.S., Pokharkar, V.G., Gavali, A.V. and Yadav, D.B. 2015. Forecasting the prices of onion in Lasalgaon and Pimpalgaon market of Western Maharashtra. International Journal of Tropical Agriculture, 33(4): 3563-3568.

IV. Damodar N Gujarati, Dawn C Porter, Sangeetha Gunasekar - Basic Econometric‖ fifth edition 2012.

V. Chengappa P.G et al., (2012). Competitive Assessment of Onion Markets in India. Competition Commission of India.

VI. GoI, (2017). Directorate of Economics and Statistics. data base report available at https:// eands.dacnet.nic.in/

VII. https://corporatefinanceinstitute.com/resources/knowledge/finance/forecasting/ (CFI)